



Express Mail Label No. EL 443 497 265 US

APPLICATION FOR LETTERS PATENT OF THE UNITED STATES

NAME OF INVENTOR(S):

Michele C. Minninger
314 Nottingham Drive
Spring City, PA 19475
UNITED STATES OF AMERICA
Citizen: USA

TITLE OF INVENTION:

A Data Transformation System

TO WHOM IT MAY CONCERN, THE FOLLOWING IS
A SPECIFICATION OF THE AFORESAID INVENTION



A DATA TRANSFORMATION SYSTEM

I. BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention provides a system and a method for transforming data of a first data structure to a different second data structure that is compatible with an executable application. In particular the present invention relates to a system and method for software conversions which maps and translates data from model reports to files 10 formatted as either Comma Separated Value (CSV) files or Flat Fixed Position files in order to migrate the data from one software systems' service to a similar service of a different software system. Using a report generated from the current system and a translation tool the data is mapped and translated to the receiving systems' conversion 15 format. By developing standard reports and pre-defined translation templates data from one system's application can be converted easily, repeatedly and efficiently.

The function of the present invention is to map and translate data from one system to another in a way that allows for the system to be delivered and executed uniformly but easily customized and that allows customers to obtain their data inexpensively and 20 repeatedly.

2. The Related Prior Art

Prior conversion methods required development efforts that were inherently 25 ridged, expensive, and labor intensive. There are multiple ways that prior conversions

were performed. Two are listed here:

One method used a COBOL program to translate and map the data from the current system to another system. This method was undesirable because every customer needed their own COBOL program and any modifications in mapping or translation needed by
5 the customer required COBOL programming modifications.

Another method used COBOL programs to load the data to be converted into a tape file. The tape file was then loaded into a spreadsheet for mapping and translation. This method posed problems do to the intervention and resulting cost needed to create the tape file and the limitations and inadequacies of the spreadsheet's ability to map and translate
10 data.

II. SUMMARY OF THE INVENTION

The present invention overcomes the aforementioned drawbacks of the prior art proposals by providing a system and a method for software conversions which maps and translates data from model reports to files formatted as either Comma Separated Value
15 (CSV) files or Flat Fixed Position files in order to migrate the data from one software systems' service to a similar service of a different software system. Using a report generated from the current system and a translation tool the data is mapped and translated to the receiving systems' conversion format. By developing standard reports and pre-defined translation templates data from one system's application can be converted easily,
20 repeatedly and efficiently. The current systems data can be obtained through standard report procedures, which is cost effective and easily repeated.

The mapping is standardized which decreases implementation time.

The translation is easily accomplished. The cost to the customer is reduced while the time

to completion is also reduced.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG.1 is a general flow chart of the present invention; and

FIG.2 is detailed flow chart illustrating step 12 of FIG.1.

5

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

Referring to the drawings and in particular FIG.1 which shows a flow chart illustrating the method and system of the present invention in which data elements from a first data structure are acquired by a preprocessor. The preprocessor collates the data 10 elements into a source file having a source format (step 11).

The first data structure has the ability to be accessed by random key and is given a specific name in the system. The data elements within the structure are identified by column, length, and type (text, money, integer, date, decimal) and given specific element names. These identifications are housed and maintained by the system. The system can 15 maintain multiple first data structures and elements where the layouts are different. To create the second data structure, the necessary pre-defined data elements are selected and given an order and spacing output format by the pre-processor. The records to be extracted have the option to be determined by data element criteria. For example, one can generate the output for all records where the data element named VENDOR has a 20 value between 0 and 9000. A specific name is given to the output format and its associated extraction criteria. The system houses and maintains multiple output specifications for each defined data structure. At any given time, the user can request the

generation of the second data structure based on the content of the current first data structure and selected output format specification.

Step 12 of FIG. 1 describes how the data elements in the source file are mapped by a processor into an output file having a second data format. As noted in step 12 this 5 mapping occurs in response to a selected predetermined control data file of a plurality of control data files that determine a corresponding plurality of different second data formats.

The second data structure is in either flat file or text format. The second data structure is accessed and the section of repeating data is identified. For example, if seven 10 lines of data represent a Vendor, then seven lines containing vendor data would be selected. Within the selected group, data elements would be identified as those needed in the mapping. Each element to be mapped is identified by row, column, and length and given a name. In addition, the data element(s) that uniquely identifies a repeating group is/are identified. The system is then able to identify by name the data elements for each 15 repeating group with in the second data structure. The Predetermined Control Data File contains these data element identifications and repeating group criteria. In addition, the mapping specifications for the output file are defined and kept in the Predetermined Control Data File. The mapping specifications contain the following data: field name, field length, field type, and field value. The field value can be a hard coded value, a data 20 element from the repeating group, or instructions for deriving the field value. A derived field would use system-defined functions to manipulate the data. System functions would include, but not be limited to, arithmetic instructions, absolute value, Boolean expressions, If-then expressions, address parsing, name-parsing, data justification, and

date reformatting. The Predetermined Control Data File is independent of the actual data in the second data structure, but is dependent on the second data structure format. When requested, the system will apply the Predetermined Control Data File against the contents of the current second data structure and produce a mapped output file. FIG. 2 illustrates 5 step 12 of FIG.1 in detail. As can be seen, a number of determinations about the Predetermined Control Data File have to be made, such as identifying a row column structure (step 12a), identifying particular data elements to be mapped (12aa), identifying source file and corresponding output file locations of particular data elements to be mapped from the source file to the output file (12c) and identifying a row column 10 structure for the source data format (step 12d).

Step 13 of FIG.1 relates to the feature of cross mapping. Sometimes the data elements from the first data format cannot be converted by the processor into the second data file format. In such cases the data elements are retained and cross mapping is done. The mapped output file is used as input along with an “old –new” file. The “old-new” 15 file is any data structure that contains the necessary data element(s) and the value it is in the old format- the first data format and the value it should be in the new format – the second data format. The system will read each record in the mapped file and change the necessary data element(s) from the “old” value to the “new” value found in the “old-new” file (step 14). Possible errors are tracked (see step 15 of FIG. 1). For example if the 20 value in the mapped file does not exist as an old value in the “old-new” file. If errors are identified, the mapped file will not be updated, but the system reports the error situations (see step 16 of FIG. 1). If no errors are identified, then the system updates the values in the mapped file and reports the changes (see step 17 of FIG. 1).

The following example will provide an illustration of the present invention's method and system in operation:

The users want to convert from a first system having a Vendor Master file and Invoice Master file data to a second system having a financial system. The second system has separate Vendor and Invoice data files. The first system's file layouts are predefined, as are the second system's file layouts. The first system's Vendor Master file is defined in the pre-processor system with the following data elements assignments:

CUSTOMER NUMBER, length 4, positions 1 – 4, type numeric;

10 VENDOR NUMBER, length 7, positions 5 – 11, type numeric;

VENDOR TYPE, length, 6, position 12 – 17, type alphabetic;

VENDOR NAME, length 30, positions 18 – 47, type alphanumeric; and

REMITTANCE ADDRESS, length 80, positions 48 – 127, type alphanumeric;

An output assignment called VENDOR LIST is created. VENDOR LIST is

15 defined to generate output for all records. The text output will be generated in the

following three-line format:

VENDOR NUMBER line 1 positions 1- 7

VENDOR TYPE line 1 positions 8 – 13

VENDOR NAME line 2, positions 1 – 30

20 REMITTANCE ADDRESS line 3, positions 1 - 80

The first system's Invoice Master file is defined in the pre-processor's system with the following data elements assignments:

CUSTOMER NUMBER, length 4, positions 1 – 4, type numeric;

VENDOR NUMBER, length 7, positions 5 – 11, type numeric;
INVOICE NUMBER, length 10, positions 12 – 21, type alphanumeric;
INVOICE TYPE, length, 6, position 22 – 27, type alphabetic;
INVOICE AMOUNT, length 10, positions 28 – 37, type money;
5 INVOICE DATE, length 8, positions 38 – 45, type date mmddccyy; and
DESCRIPTION, length 20, positions 46 – 65, type alphanumeric.

An output assignment called INVOICE LIST is created. INVOICE LIST is defined to generate output for all records with an INVOICE TYPE of “ACTIVE”. The text output will be generated in the following two-line format:

10 VENDOR NUMBER line 1 positions 1- 7
INVOICE NUMBER line 1 positions 8 – 17
VENDOR TYPE line 1 positions 18 – 23
INVOICE DESCRIPTION line 2, positions 1 – 20
INVOICE DATE line 2, positions 21 - 28
15 INVOICE AMOUNT line 2, positions 29 - 38

The VENDOR LIST output is accessed by the generic Mapping system. A Control Data file called, VENDOR CONTROL is setup to recognize that three lines represent one vendor and to identify the data element layout of VENDOR LIST. Next the Control Data file is updated to include the mapped output file layout as follows in Table 1:

POSITIONS	NAME	TYPE	ELEMENT VALUE
1 – 4	Company	Numeric	“1234”
5 – 14	Vendor	Numeric	Sequential starting at 1000 incrementing by 1.
15 – 44	Vendor Name	Alphanumeric	Equal to VENDOR NAME on VENDOR LIST
45 – 45	Vendor Status	Alphabetic	If VENDOR TYPE on VENDOR LIST is equal to “active”, then “A”. If VENDOR TYPE is equal to “inactive”, then “I”.

5 Table 1

The VENDOR LIST output is accessed by the present invention's Mapping system.

Another Control Data file called, CROSS FILE CONTROL is setup to recognize that three lines represent one vendor and to identify the data element layout of VENDOR

LIST. Next the CROSS FILE CONTROL Control Data file is updated to include the mapped output file layout as follows in Table 2:

POSITIONS	NAME	TYPE	ELEMENT VALUE
1 – 4	Company	Numeric	“1234”
5 – 11	Old Vendor	Numeric	Equal to VENDOR NUMBER on VENDOR LIST
12 – 21	New Vendor	Numeric	Sequential starting at 1000 incrementing by 1.

Table 2

The INVOICE LIST output is accessed by the present invention's Mapping system. A Control Data file called, INVOICE CONTROL is setup to recognize that two lines represent one invoice and to identify the data element layout of INVOICE LIST. Next the Control Data file, INVOICE CONTROL is updated to include the mapped output file 5 layout as follows in Table 3:

POSITIONS	NAME	TYPE	ELEMENT VALUE
1 – 4	Company	Numeric	“1234”
5 – 14	Vendor Number	Numeric	Equal to VENDOR NUMBER on INVOICE LIST
15 – 24	Invoice number	Alphanumeric	Equal to INVOICE NUMBER on INVOICE LIST
25 – 44	Description	Alphabetic	Equal to INVOICE DESCRIPTION on INVOICE LIST
45 – 50	Date	MMDDYY	Equal to INVOICE DATE on INVOICE LIST
51 – 60	AMOUNT	MONEY	Equal to INVOICE AMOUNT on INVOICE LIST

TABLE 3

The Mapping system will extract data from the VENDOR LIST using the VENDOR CONTROL and call it MAPPED VENDORS.

The Mapping system will extract data from the VENDOR LIST using the CROSS FILE CONTROL and call it CROSSFILE MAP.

- 5 The Mapping system will extract data from the INVOICE LIST using the INVOICE CONTROL and call it MAPPED INVOICES.

The Cross File Mapping system will access each record in MAPPED INVOICES. It will look for a match between the invoice number and the old invoice value on the CROSSFILE MAP. If all MAPPED INVOICE invoice numbers have corresponding

- 10 CROSSFILE MAP old invoice numbers, the system will change the MAPPED INVOICES invoice numbers to be equal to the corresponding CROSSFILE MAP new invoice numbers. If a MAPPED INVOICES invoice number does not exist as a CROSSFILE MAP old invoice number, no changes are made and an error report is generated.

- 15 Assuming that the cross file mapping was successful, MAPPED VENDORS and MAPPED INVOICES are now available for input in to the second system's financial system.

Although the present invention has been described in terms of the presently preferred embodiments, it is to be understood that the disclosure is not to be interpreted 20 as limiting. Various alterations and modifications will no doubt become apparent to those skilled in the art after having read the above disclosure. Accordingly, it is intended that the appended claims be interpreted as covering all alterations and modifications as fall within the true spirit and scope of the invention.